Journal of Digital Health
https://ojs.luminescience.cn/JDH

LUMINESCIENCE
https://luminescience.cn/

Original Research

AI-based big data analysis of user-generated content to understand adherence in oral hormone receptor-positive advanced breast cancer treatment: A literature review and patient-centered approach using natural language processing

Stefanie Scholz [1*], Christian Winkler [2], Svenja Niescken [3]

[1] SRH Wilhelm Loehe University of Applied Sciences, Merkurstrasse 19, 90763 Fuerth, Bavaria, Germany
[2] Nuremberg Institute of Technology – University of Applied Sciences, Kesslerplatz 12, 90489 Nuremberg, Germany
[3] Deputy Managing Director, Informationsdienst Wissenschaft e.V., Bayreuth, Germany

*Correspondence: Stefanie Scholz, Email: Stefanie.scholz@srh.de

Abstract:
Background: Understanding patients' adherence needs is crucial for the success of therapies, particularly for chronic illnesses and cancer. The shift from intravenous to oral anticancer medication enhances patient independence but may threaten adherence. This study aims to identify patients' concerns and needs for treatment of oral hormone receptor-positive for advanced breast cancer (CDK4/6 inhibitors).
Methodology: A comprehensive literature review on factors influencing patient adherence was conducted, followed by AI-based big data analyses of user-generated content. Techniques such as natural language processing and machine learning-based word embeddings were applied to achieve a patient-centered understanding of adherence.
Results: The literature review shows that adherence is often overlooked in clinical research due to the impracticality of long-term monitoring in most trial designs. This highlights the need for alternative methods such as real-world data and AI-based analyses to understand patients' long-term adherence to treatment. Key factors influencing adherence to oral breast cancer treatment include the cost-benefit ratio in terms of survival and quality of life, realistic expectations regarding side effects, social support and personalized care. AI analysis of user-generated content supports these findings and reveals more frequent discussions on non-adherence in early breast cancer stages. Semantic graphs and topic modeling also identified additional patient concerns and informational needs for better patient communication.
Conclusions: Enhancing patient-centeredness is crucial for improving adherence and persistence to oral treatment for breast cancer. The increased autonomy provided by oral medications offers a significant opportunity for healthcare providers to empower patients in managing their treatment. Comprehensive support, including addressing emotional and psychological issues, is essential, especially in advanced breast cancer stages.

Keywords: Adherence, Patient centricity, Breast cancer, Big data, Natural language processing, User-generated content

Received: Jun. 4, 2024; Revised: Aug.1, 2024; Accepted: Aug.8, 2024, Published: Aug.13, 2024

DOI: https://doi.org/10.55976/jdh.32024126917-34

Background

Over the last decades, the increasing use of targeted oral anticancer medications (OAMs) has significantly transformed cancer patient management. This shift has moved from directly observed, intermittent intravenous therapy to self-administered, oral chronic therapy [1-3]. The steady rise in OAM use has offered numerous patient benefits but also poses a notable challenge for healthcare professionals (HCP) in accurately monitoring patient adherence. Nonadherence is linked to poor health outcomes and decreased survival [4-6]. Consequently, medication adherence has become a critical focus in oncology treatment due to the rise of targeted OAMs. Given the fact that the incidence of breast cancer (BC) is increasing, making it the most prevalent cancer worldwide [7].

Introduction

It is essential to understand patients' concerns and needs regarding their treatment, particularly for treatment of oral hormone receptor-positive for advanced breast cancer (CDK4/6 inhibitors) [8, 9].

Based on the literature, there are several identified factors that influence patient adherence to oral anticancer therapies, particularly in the context of metastatic breast cancer. These factors can be broadly categorised into predisposing, treatment-related and healthcare system-related factors (as shown in Figure 1).

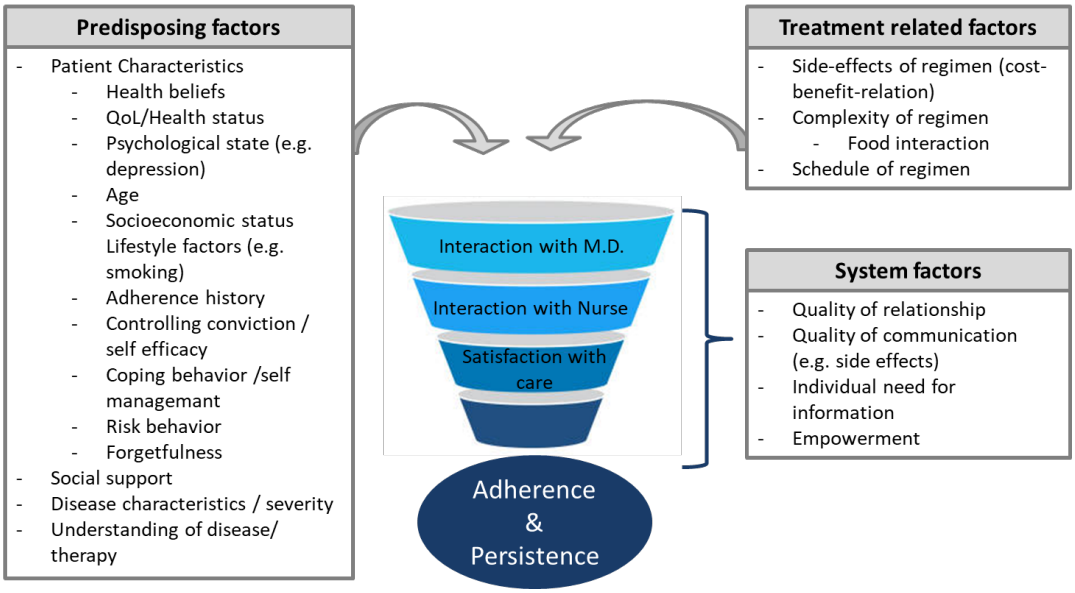

Figure 1. Factors influencing adherence to OAM (based on 41)

Predisposing factors include patient-related factors such as socio-demographics such as age, gender, race, education level, income and employment status [10]. Studies have shown that younger age, higher levels of education and employment are associated with better adherence to oral anticancer therapies [11, 12]. Additionally, psychological factors such as anxiety, depression and health-related quality of life have also been found to influence treatment adherence [13-19]. For example, patients with high levels of anxiety and depression may be less likely to adhere to their medication regimen [20-23].

Treatment related factors include the complexity of the medication regimen, side effects and medication cost. Studies have shown that medication complexity and side effects are major barriers to adherence to oral anticancer therapies [24-28]. Similarly, medication cost can also be a significant barrier to adherence, particularly for patients who have limited financial resources [29].

Factors related to the healthcare system include access to healthcare, communication with healthcare providers and the quality of healthcare services. Patients who have better access to healthcare and who have good communication with their healthcare providers are more likely to adhere to their medication regimen [17, 30, 31]. Additionally, the quality of healthcare services, such as patient education and support, also plays an important role in treatment adherence [32-35].

In the context of metastatic breast cancer, adherence to oral anticancer therapies is particularly important, as these therapies are often the mainstay of treatment for these patients [36-51]. Nonadherence to these therapies can lead

to disease progression, reduced survival and increased healthcare costs [49, 52-55]. Therefore, it is essential to identify the factors that influence patient adherence to treatment in this population and to develop interventions to address these factors [56-62].

In summary, patient-related, medication-related and healthcare system-related factors can all influence patient adherence to oral anticancer therapies. Both, direct and indirect measurement of adherence remains a challenge, as it can be insufficient asking HCP how they assess their patients' adherence or be biased through social desirability or inhibitions regarding sensitive topics [63].

In the context of metastatic breast cancer, adherence to these therapies is critical, and understanding the factors that influence adherence is essential for optimizing patient outcomes [64].

In this study, we aimed to identify patients' specific adherence needs and issues by conducting a literature review and analyzing user-generated content (UGC) provided by patients using AI-based big data analyses. We sought to provide a patient-centered understanding of adherence and highlight the key aspects that influence patients' adherence to oral breast cancer treatment. Additionally, we explored the patient-centered (information) needs, concerns and related issues that should be considered for a patient-oriented communication. Our findings emphasize the importance of patient centricity and social and professional support, not only regarding "hard facts", but also emotional and psychological aspects to strengthen the support of patients.

Traditional clinical studies often focus on the clinical efficacy and side effects of treatments and pay limited attention to patient adherence, particularly for oral medications, as the long- term monitoring is not feasible in most clinical trials or study designs [65, 66]. By integrating a literature review with AI-based analysis of UGC, our approach captures a broader spectrum of patient experiences and concerns. Leveraging real-world data from UGC provides a potential for valuable insights about genuine patient experiences and perspectives. Analyzing this data using AI allows us to uncover nuanced insights into patient behavior, adherence challenges, and the emotional and psychological dimensions of managing oral cancer treatments. This approach emphasizes understanding adherence from the patients' perspective. By focusing on patient-generated content, we can identify specific needs, concerns and factors that influence adherence, thereby ensuring that our findings are relevant and applicable to patient care. Understanding the patient perspective can help providers develop strategies that address not only the clinical aspects of treatment, but also the emotional and psychological needs of patients, ultimately improving adherence and outcomes.

To take the specific needs of patient journey into account when analyzing the patient perspectives on adherence-related concerns and thoughts we concentrated on user-generated content from patients with metastatic breast cancer.

User-generated content

As our analysis depends on using suitable content, considerable effort is spent in identifying the most relevant websites. This process makes heavy uses of search engines and in this way mimics the behavior of users.

As we do not know what users are searching for, we start with keywords in different categories – in our case, we used the categories symptoms, medication and adherence. For each category, we started with just a few keywords and organically expanded them with semantically similar words. To do this, we used word embeddings (trained from the corpus itself), search suggestions and semantic graphs. Initially, a small set of foundational keywords was selected for each category. These keywords were broad and representative of the main concepts within each category. For instance, in the "symptoms" category, the keywords included specific symptoms associated with the condition or treatment. Similarly, in the "medication" category, the keywords could encompass drug names, types of treatments, or related pharmaceutical terms. For "adherence," the keywords may involve terms related to compliance, patient behavior, or factors influencing adherence. This approach helped to uncover synonyms, related concepts and variations that users or patients might employ in their searches but may not have been initially considered. Search suggestions, often derived from search engines, provide insights into commonly searched terms and are based on actual related queries. These suggestions were used to validate the initial selection of keywords and revealed additional keywords that aligned with users' search behaviors. Semantic graphs visualized the relationships between words or concepts based on their semantic proximity or co-occurrence in texts. Analyzing such graphs helped us to identify clusters of related terms and enabled the inclusion of comprehensive keyword sets that covered various aspects of the study's focus areas. Through iterative refinement using these methodologies, the initial set of keywords within each category was expanded to approximately 20 relevant terms per category (as shown in Table 1 below). This iterative process ensured that the keyword selection was comprehensive and covered a diverse range of user search intentions. This enhanced the study's ability to attract relevant user traffic or inform further research and communication strategies. This approach to identifying keywords integrated both initial selection and iterative expansion using semantic analysis tools to ensure that the study could effectively address users' search queries and information needs related to symptoms, medications and adherence in the context of the research.

Table 1. Search keywords in categories.

Symptoms	Medication	Adherence
mammary cancer	oral anticancer agents	patient outcome
mammary neoplasm	oral cancer agents	medication compliance
mammary carcinoma	oral anticancer drugs	medication persistence
mammary tumor	endocrine therapy	medication adherence
breast cancer	oral endocrine therapy	medication nonadherence
breast neoplasm	aromatase inhibitor	compliance
breast carcinoma	ribociclib	persistence
breast tumor	Kisqali	adherence
metastatic breast cancer	Alpelisib	nonadherence
metastatic breast neoplasm	Piqray	compliant
metastatic breast carcinoma	Everolimus	adherent
metastatic breast tumor	Afinitor	comply
advanced breast cancer	Ibrance	adhere
advanced breast neoplasm	Verzenio	take my medicine
advanced breast carcinoma	Verzenios	take medicine
advanced breast tumor	Palbociclib	skip my medicine
	Abemaciclib	doctor
	Letrozol	physician
	Anastrozol	oncologist
	Exemestan	communication
	Fulvestrant	pill reminder
	Tamoxifen	forget pill
	Femara	
	Arimidex	
	Aromasin	
	Faslodex	
	Nolvadex	

Building all possible combinations results in 10,295 search requests (we omitted searches for single words in the adherence category as these turned out to be too unspecific). Search engines such as Google and DuckDuckGo can be configured to yield up to 30 search results per requests, leading to 524,835 search results.

These search results can be analyzed in different dimensions. We are primarily interested in finding the websites carrying the most content. Therefore, we aggregate the results with respect to the website and count how often they appear in the top 30 results (see Table 2). This list must then be manually classified whether the websites contain editorial or user-generated content. We are interested in the latter.

It turns out that a website with much content is www.breastcancer.org. Fortunately, the editorial content is complimented by user-generated content in the form of a forum called "forum.breastcancer.org". The website makes a professional impression and is frequented visited by many users. Therefore, it is an ideal candidate for acquiring user-generated content. For this reason, a custom URL-generator, a downloader and an extractor were implemented.

The breastcancer.org forum is large with more than 5 million posts. Fortunately, it is organized in different subforums where patients with different stages of cancer write posts and comments. As we are interested in comparing the different stages, there is no need to implement an additional classifier for discriminating between the different stages (see Table 3 for the structure and number of posts for each subforum of forum.breastcancer.org). All extracted content is stored in a SQLite database to make later retrieval as easy as possible.

Content analysis

UGC was analyzed by employing AI-enhanced methodologies. The dataset utilized for this investigation was sourced from publicly accessible forums, identified through thorough desk research. It is crucial to underline that the approach taken prioritized data privacy protection. Individual user comments or posts remained untouched or unexamined, adhering to stringent privacy protocols. All findings presented herein stem from a comprehensive aggregation of statements, performed using AI-driven

Table 2. Top 20 websites in search results

Domain	Count
www.ncbi.nlm.nih.gov	20426
www.drugs.com	14344
www.breastcancer.org	13575
www.researchgate.net	11439
ascopubs.org	11354
pubmed.ncbi.nlm.nih.gov	10987
www.sciencedirect.com	9900
www.webmd.com	9677
clinicaltrials.gov	9564
www.cancer.gov	8833
ww5.komen.org	8323
www.cancer.org	8176
www.annalsofoncology.org	7954
link.springer.com	6313
www.healthline.com	5977
www.nature.com	5445
www.mayoclinic.org	5332
en.wikipedia.org	5011
www.cancer.net	4727
www.karger.com	4655

Table 3. Top forums of breastcancer.org

Forum	Number of posts
Forum: Humor and Games	653272
Forum: Stage IV/Metastatic Breast Cancer ONLY	541737
Forum: Chemotherapy - Before, During, and After	523223
Forum: Breast Reconstruction	300201
Forum: Bonded by Breast Cancer	212370
Forum: Surgery - Before, During, and After	209928
Forum: Life After Breast Cancer	179178
Forum: Hormonal Therapy - Before, During, and After	166491
Forum: Mental Health: Because Cancer Doesn't Just Affect Your Breasts	158969
Forum: Not Diagnosed But Worried	139243
Forum: Radiation Therapy - Before, During, and After	134968
Forum: Just Diagnosed	105846
Forum: Prayers and Spiritual Support	84915
Forum: Working on Your Fitness	82770
Forum: Triple-Negative Breast Cancer	78670
Forum: HER2+ (Positive) Breast Cancer	74763
Forum: Lymphedema	73381
Forum: Older Than 60 Years Old With Breast Cancer	72807
Forum: Managing Side Effects of Breast Cancer and Its Treatment	72797
Forum: IDC (Invasive Ductal Carcinoma)	67834

techniques, primarily leveraging natural language processing (NLP).

When examining the aspects related to adherence, it was essential to consider that individuals within the patient community do not typically employ terms such as "adherence" or comparable expressions when discussing their treatment behaviors and concerns. To overcome this, semantic techniques were applied to identify elements associated with the notion of "adherence." Semantic transformation is the manipulation of text data to change its representation while preserving or improving its underlying meaning. It is essential for various NLP tasks, such as information retrieval, sentiment analysis, machine translation and text summarization. The primary objective of semantic transformation is to bridge the gap between the intricacies of human language and machine comprehension and enable computers to work more effectively with textual information. One of the

basic techniques for semantic transformation is word embeddings such as Word2Vec, GloVe or fastText. These methods convert words or phrases into high-dimensional vectors, positioning them in a semantic space where similar meanings are closer to each other. This transformation captures semantic relationships that allow algorithms to understand word context and similarity. Our corpus was large enough to train a dedicated word embedding model.

To get a semantic vector representation of a post, we calculated the TF/IDF representation of this post within the whole corpus. We used the weights of the individual tokens as prefix factors for the word embeddings and calculated a TF/IDF-averaged total embedding te:

$$\mathbf{te}\ (\text{Post}) = \sum_{w \in \text{Post}} \text{weight}_w \cdot \mathbf{wv}(w)$$

Compared to semantic methods such as large language models, this approach is conceptually simpler, but lets us evaluate the similarity of a post to a single word w which is represented by the cosine distance of the word vector wv (w) and te (Post):

$$\text{sim}(\text{Post}, w) = \cos \angle\big(\mathbf{te}(\text{Post},), \mathbf{wv}(w)\big) = \frac{\mathbf{te}(\text{Post}) \cdot \mathbf{wv}(w)}{\|\mathbf{te}(\text{Post})\| \, \|\mathbf{wv}(w)\|}$$

The data analysis was carried out in a series of meticulously orchestrated steps, each of which contributed to unveiling the comprehensive insights. The subsequent sections illuminate the intricacies of this analytical journey and provide a clear and structured depiction of the methodologies employed.

Topic model methodology

To understand the structure of a large number of documents (a corpus), topic models [67] can be used. This approach has become very popular in the last decade [68]. The methodology can uncover the latent structure of the corpus and identify topics. The main interest is the structure of these topics rather than the individual assignment of documents to topics. If the main interest were in the latter, clustering would be a more appropriate method.

This unsupervised method works very well and can be tuned in various dimensions [69]. The starting point is the document-term matrix, which counts how often each word occurs in each document. The vocabulary is enumerated and each word is assigned a number (e.g., in the order of occurrence). Stop words, i.e., words that carry no meaning, can already be eliminated in this step. The order of words is lost. This is a serious restriction and can be remedied by considering two-word combinations (bigrams). The document-term matrix contains the words in the columns and the documents in the rows. Its dimensions are (number of documents) × (number of different words). Each cell contains the number of times the word occurs in the document. Sometimes, a TF/IDF (Term Frequency-Inverse Document Frequency) transformation is applied to penalize words that occur very often in the corpus [70]. The TF/IDF matrix does not contain any information about the author and ensures that the content is properly anonymized. As each post is as important as any other, the analysis based on the top of the matrix is unbiased.

The matrix can become very large and is not suitable for direct interpretation. However, the matrix only contains non-negative values. Such matrices can be decomposed into a product of smaller matrices with k rows/columns, where k is the rank of the non-negative matrix. An approximation can be made by choosing a (much) lower k than the actual rank, which makes both factor matrices interpretable. The matrix H in Figure 2 is the document-topic matrix and shows the topic contributions to each document. W is called the topic-term matrix and contains the word contributions to each topic: $\mathbf{D} \approx \mathbf{W} \times \mathbf{H}$

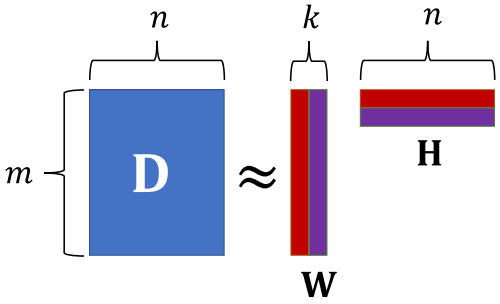

Figure 2. Decomposing a document-term matrix into a document-topic and topic-term matrix (own illustration).

In topic modelling, the document-term matrix is usually transposed, meaning that it consists of n documents and m terms. This method is applied to the corpus in the next section.

Creation of topic models

In order to qualitatively examine the general topics of the online community, a topic modelling was first performed. Since the topics of the individual sources usually do not match, the topic modelling must be carried out separately for each data source. In order to enable subsequent interpretation of the data, the number of topics per website was limited to ten. Word clouds of the topics were provided for qualitative visual illustration. Stopwords were already eliminated in advance for all websites (e.g.

"deleted", "removed", "thanks", "thank", "sharing", "user", "https", "www" and "com"). After obtaining the first topic models, it was necessary to eliminate further stopwords, as they had too little semantic value or were not relevant to the analysis of the research question.

Analysis of the topic models

For each topic, subtopics can be calculated and show words/topics that occur frequently in this context. Furthermore, the topic sizes show the individual relevance per source for the respective topics. The relevance of each word in the topic is (qualitatively) indicated by the size, but can also be represented by a number. For each source, both unigrams, consisting of individual elements, and two-word combinations, so-called bigrams, were created. The data can be saved in Excel format for further analysis.

In order to create a basis for the interpretation that follows in the further course of the work, the topics of the individual sources were sorted according to their relevance. The lower the weighting, the less relevant the topic was within the present topic models. The relevance of the individual topics was expressed by numbers: the number 0 expresses a high relevance, while the number 9 represents the lowest relevance within the present topic models.

→ 0 = very relevant, 4 = relevant, 9 = less relevant.

The weighting of the subtopics (terms within a topic) was considered in addition to the topics in order to assign them to the categories. Topics with relevance scores from zero to four could be considered for category assignment. Finally, the terms within the topics were examined more closely for feelings or emotions. The terms were assigned to a more positive or a rather negative expression.

Topic evolution analysis

For the analysis of the time development of the topic models, the size of each individual topic can be calculated for specific time periods (e.g., months). The result can be interpreted as a time series and visualized in a diagram. The data was also exported to Excel format for subsequent calculations.

Semantic transformation

Semantic transformation in the field of NLP refers to the advanced technique of modifying the meaning or structure of a text while retaining its underlying semantic essence [71]. This transformation process is a key analytical approach employed alongside other methodologies in NLP research [72]. The primary aim of semantic transformation is to generate a text that maintains the core meaning of the original content but may be articulated differently or convey a distinct sentiment, emphasis or style. In our study, we used semantic transformation to identify discussions dealing with adherence-related aspects, but were not mentioned or used terms such as adherence, compliance, etc. By applying semantic transformation, sensitive information can be obscured while preserving the overall message. This is particularly useful for sensitive research topics such as in our study, where analysing data requires privacy protection. Using semantic transformation, we could broaden the scope of NLP applications and enhance language understanding.

Description of data basis: community.breastcancer.org

Utilizing data spanning from January 2001 to December 2022, approximately 4.7 million posts from the breastcancer.org forum were analyzed. Figure 3 illustrates the continued vibrant engagement of active users, which is reflected in the quarterly accumulation of around 5,000 posts by users within the dedicated subforum "Stage IV/ Metastatic Breast Cancer ONLY."

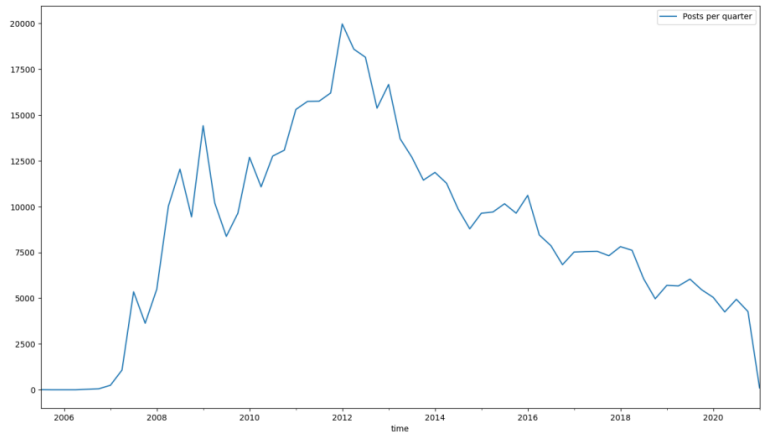

Figure 3. High number of posts per quarter in forum "Stage IV/Metastatic Breast Cancer ONLY", despite slightly decreasing number of posts since 2012

Patient journey-related most frequent bigrams: Stage IV (ONLY)

The visualization of the word cloud of prevalent bigrams within the forum dedicated to individuals with metastatic breast cancer reveals distinct areas of emphasis. Notably, the most frequent combinations of two words underscore specific focal points. These encompass a comprehensive exploration of aspects related to hormonal therapy, including notable mentions of specific CDK4/6 inhibitors, namely "ibrance palbociclib" (as shown in Figure 4).

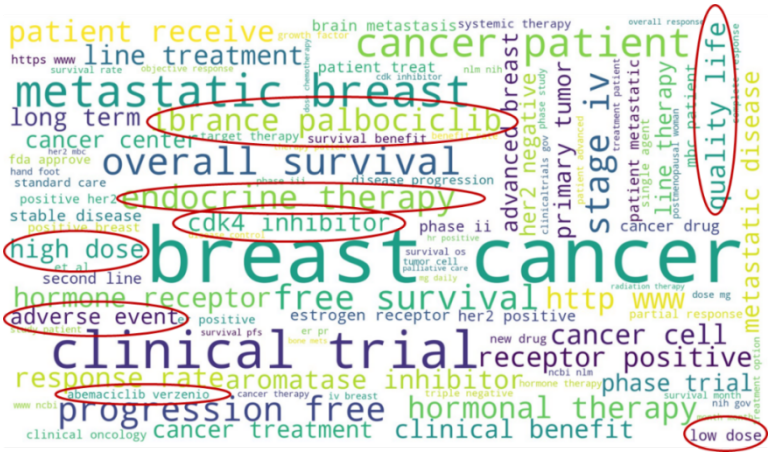

Figure 4. Most frequent Bigrams within the subforum dedicated to discussions in the realm of metastatic breast cancer (own illustration)

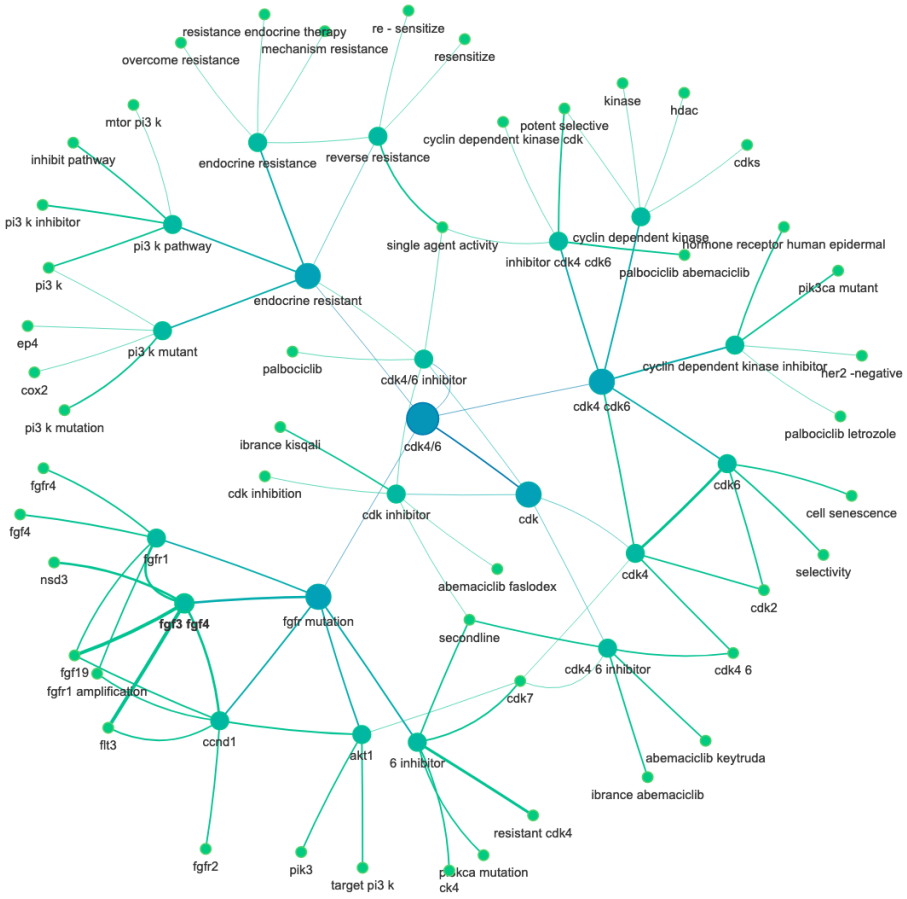

Figure 5. Semantic relationships of words to CDK 4/6. Words farther away are less related.

The occurrences of "abemaciclib verzenio" are less frequent, with discussion extending to considerations of dosage. Another salient theme that emerges prominently is the discourse on "quality of life", which serves as a recurring and extensively deliberated subject matter within the forum's conversations. In order to exploratorily dive deeper into the discussions related to adherence-specific topics, a semantic graph visualization was used.

Semantic map CDK 4/6 inhibitors: community.breastcancer.org

A semantic graph visualization portrays the interconnectedness and semantic relationships between different concepts or entities. It represents these elements as nodes, while their relationships are depicted as edges connecting the nodes. This visual representation offers an intuitive and accessible way to explore and comprehend the semantic similarities and associations within a dataset. By revealing the intricate web of connections, a semantic graph enhances the understanding of how words or concepts relate to one another, facilitating information retrieval, knowledge representation and exploratory analysis in various domains. The semantic graph, as illustrated in Figure 5, portrays CDK4/6 as a prominent focal point within the discussions, visually represented by a sizable central node.

This prominence signifies the presence of directly associated dialogues concerning endocrine resistance or resistance mechanisms and FGFR (3/4) mutations. The issue of endocrine resistance and its management is of paramount significance, as evidenced by phrases such as "overcome resistance", "reverse resistance" and "resistant CDK4", recurring in the discourse. Furthermore, the graph reveals numerous discussions that delve into the specifics of CDK4 and CDK6, indicating a high level of expertise among participants. Patients demonstrate remarkable acumen by employing precise terminology in their discussions and skillfully interconnecting related topics, as indicated by the peripheral nodes and edges. Notably, the discussions do not revolve around side effects, but instead focus on active substances and specific agents.

Discussions about "side effects" and "CDK4/6 inhibitors"

When exploratively examining the dataset through semantic graph analysis, specifically by searching for the semantic association between "side effects" and "CDK4/6 inhibitors", several prevalent terms related to the side effects of CDK4/6 inhibitors emerge. These include:

- Fatigue
- Pain/ache
- Nausea
- Diarrhea

Additionally, the analysis reveals that the instances of specific medication combinations are also mentioned. Furthermore, the discourse within the community revolves around discussions related to coping strategies, such as the concepts of tolerance, tolerability, modification or discontinuation.

Adherence-related discussions

As described above, semantic transformation was used to identify core aspects related to adherence within the community discussions. As shown in Figure 6, the semantic analyses of adherence-related aspects revealed several noteworthy findings. Firstly, there was a notable degree of semantic similarity between "adherence" and discussions pertaining to its impacts on menstrual function, FSH and GnRH-agonist. These topics often emerged in the context of These topics often emerged in the context of hormonal therapy in the adjuvant stages of breast cancer, indicating a strong thematic link. Furthermore, discussions about adherence were not limited to the patient experiences, but also extended to scientific and methodological conversations within the community, underscoring the multifaceted nature of this crucial healthcare concept.

Patient journey-related posts with adherence-content

We performed a semantic transformation and a distance analysis to determine the proportion of discussions showing significant similarity to the concept of adherence (with a threshold set at 0.4).

A notable finding can be observed that the proportion of discussions centered around adherence decreased as the patient journey progressed. This indicates a shift in the focus of these discussions as patients progressed through their healthcare experiences (see Figure 7).

Similarity of CDK 4/6 inhibitors-related discussions to adherence

To illustrate the link between discussions related to adherence and specific CDK 4/6 inhibitors, Figure 8 employs a violin plot that combines the attributes of box plots and kernel density plots. This visual representation in Figure 8 establishes a connection between adherence-related discussions and distinct CDK 4/6 inhibitors and reveals that the affinity to the adherence concept is more pronounced for the "younger" CDK 4/6 inhibitors,

Ribociclib and Abemaciclib. Conversely, Palbociclib displays a higher degree of variability in its similarity to discussions pertaining to adherence, indicating a broader range of engagement with this particular aspect.

Topic models

Topic models were calculated for six topics, as can be seen in Figure 9: Standard topic model. A qualitative visualization was used, which illustrates the relative size of the words as a measure of their importance.

The topics are constituted of many general words. Although this facilitates interpretation in terms of communication between the different participants, it misses the original goal of finding specific words and niche interests. A "happy birthday" topic is interesting and underlines the positive, empathetic atmosphere in the discussion forum, but provides little scientific insight. The same is true for "great news" and "sorry prayer".

Figure 6. Semantic similarities of aspects related to the semantic concept "adherence".

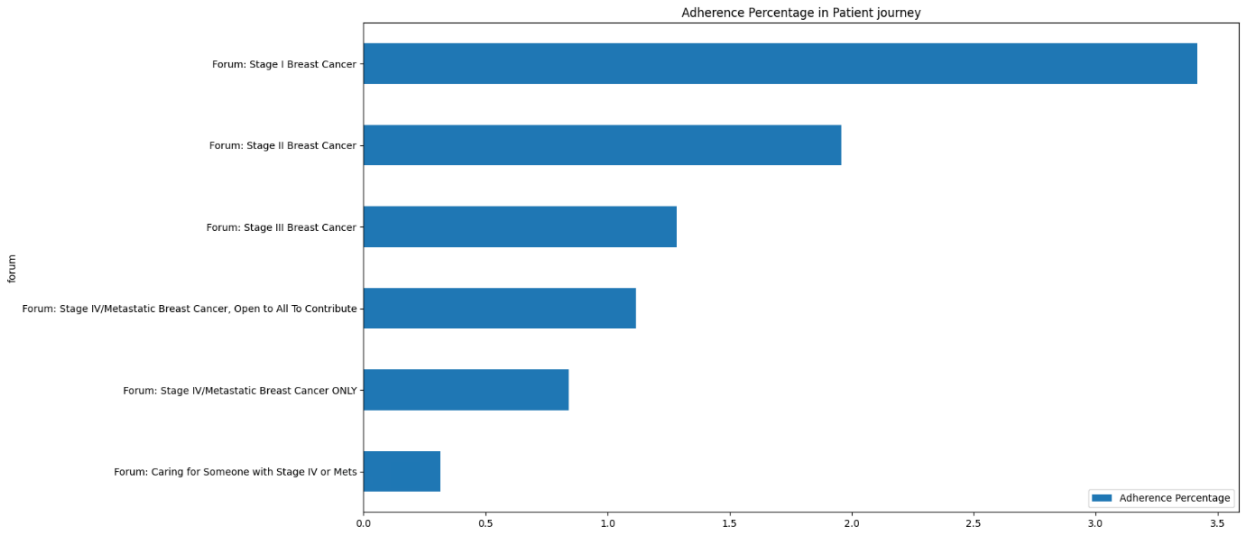

Figure 7. Share of discussions including adherence-related aspects along the patient journey.

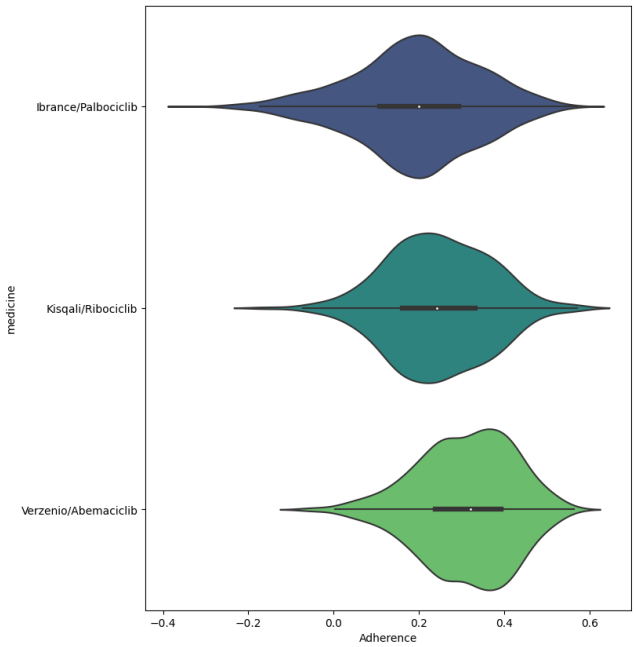

Figure 8. Violin plot illustrating the semantic similarity of discussions pertaining to the semantic concept of adherence in the context of CDK 4/6 inhibitors.

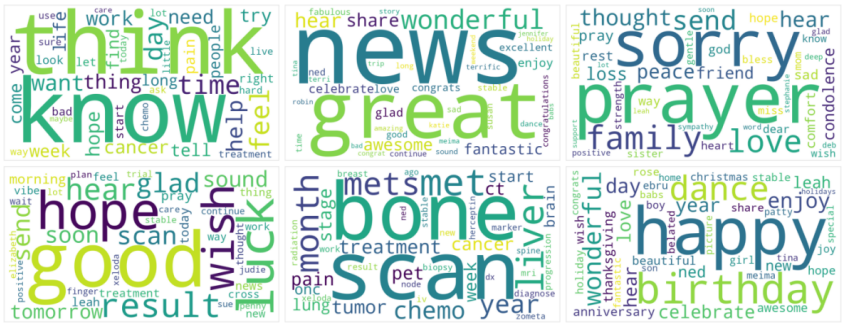

Figure 9. Standard topic model

This can be remedied by modifying the vectorization procedure and avoiding words that are too common. Using a list of the 1,000 most uncommon English words, we get the topics visualized in Figure 10: Uncommon words topic model. These topics are much easier to interpret in a scientific context.

- The first topic "chemo" contains a discussion about general chemotherapy and the typical agents used. Furthermore, the words constituting that topic are listed in Table 4: Constitutent words of "chemo" topic.
- The "onc" topic is focused on detection and progression of cancer.
- The "mets" topic discusses the diagnosis process itself with the oncologist and possible metastases.
- The topics "xeloda", "ned" and "ibrance" discuss specific medications and their effects.

Semantic map "support"

Analyzing the outcomes of the semantic graph focusing on "support" to uncover pertinent facets concerning the requirements of patients for achieving a sense of support, the subsequent findings emerge as noteworthy:

Figure 10. Uncommon words topic model

Table 4. Constituent words of "chemo" topic

Words	Proportion
chemo	25,99%
taxol	1,58%
hormonal	1,28%
oncologist	1,15%
herceptin	1,13%
abraxane	0,92%
gemzar	0,87%
diagnose	0,86%
taxotere	0,81%
rad	0,79%
infusion	0,78%
tamoxifen	0,65%
progression	0,64%
navelbine	0,62%
avastin	0,60%
zometa	0,59%
shrink	0,57%
biopsy	0,49%
nausea	0,47%
mastectomy	0,47%
halaven	0,44%
her2	0,43%
neuropathy	0,42%
fatigue	0,41%

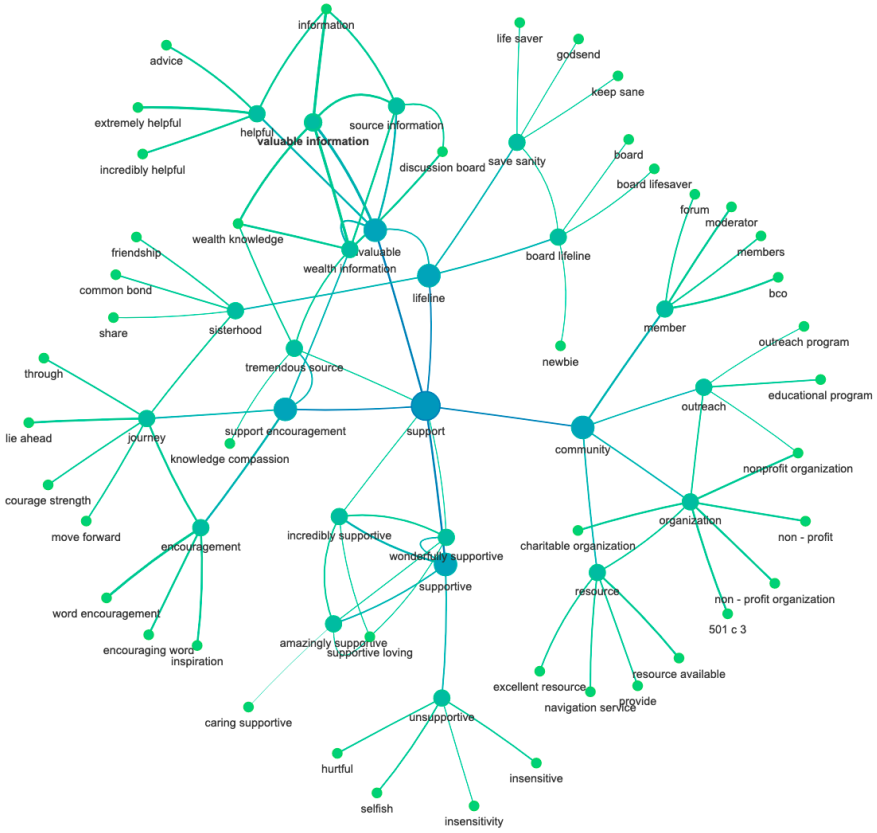

Figure 11. Semantic relationships of words to "support". Words farther away are less related.

The term "support" exhibits a robust correlation with the notion of encouragement within the community. "Support" is also perceived as a crucial source of valuable information, which includes advice and extends to encompass emotional support in the form of friendship and sisterhood as well as motivational encouragement. The concept of support is intricately linked to the idea of a journey undertaken with companions who provide assistance and encouragement. Furthermore, the concept of support is also intertwined with educational initiatives and outreach programs, suggesting a multifaceted approach to addressing patient needs.

These findings shed light on the multifaceted nature of support in relation to patient well-being, which encompasses not only informational and emotional aspects, but also collaborative and educational dimensions.

Conclusion

This study delved into the crucial realm of patients' adherence to treatment of oral hormone receptor-positive for advanced breast cancer, specifically focusing on CDK4/6 inhibitors. Through a comprehensive exploration of both the clinical literature and AI-driven analyses of UGC, the multifaceted factors influencing patients' adherence to treatment were elucidated. The dearth of attention dedicated to patients' adherence in clinical research was illuminated, highlighting the need for a paradigm shift towards patient-centricity in treatment approaches.

The findings underscored that the patient is at the center of enhancing adherence and persistence, particularly in the context of oral breast cancer treatment. The advent of oral agents offers patients increased autonomy and presents HCPs with a remarkable opportunity to empower patients in managing their cancer treatment. Moreover, it became evident that while the transition to oral therapies promotes flexibility, it also necessitates a stronger emphasis on cultivating effective communication, emotional support and psychological well-being.

The study revealed several key determinants of patients' adherence, including a positive cost-benefit balance concerning overall survival and quality of life, realistic expectations regarding side effects, and personalized support from HCPs. These determinants were consistently echoed in both the literature review and the AI-derived analyses of UGC. Additionally, the exploration of non-adherence trends across different stages of breast cancer highlighted the shifting dynamics of patient concerns and needs. Ultimately, the insights gained from this study emphasize the imperative of embracing patient-

centered care in oncology, particularly in advanced stages of breast cancer. Beyond the realm of clinical efficacy, patient empowerment, emotional support and effective communication stand as pillars for reinforcing patients' resolve in their cancer treatment journey. As medical practice continues to evolve, this research calls for a holistic approach that recognizes and addresses the intricate interplay between medical interventions and patients' individual experiences, ultimately fostering a more resilient and engaged patient population.

In a broader context, this study highlights the potential of data science and NLP to shed light on patient perspectives and support the development of targeted interventions that can bridge the gap between medical research and patient needs. As the field of oncology advances, these insights are poised to drive transformative shifts towards more patient-centered and effective treatment paradigms.

Managerial implications

HCPs should prioritize patient-centered communication by tailoring discussions to individual preferences and addressing concerns about treatment. In addition to medical care, providing a comprehensively emotional and psychological support is essential. The integration of technologies for real-time adherence monitoring can help identify and manage non-adherence promptly. Pharmaceutical companies should collaborate with HCPs to create patient-focused educational materials that emphasize the benefits of treatment and address common worries. By developing patient support programs, a concrete practical guidance on managing side effects of treatment can be fostered. The investment in research that focus on the patient's perspectives will guide the development of patient-oriented treatments and services. Healthcare policy makers could incentivize patient-centered care by implementing policies that reward healthcare systems prioritizing comprehensive patient support, shared decision-making, and effective communication. Funding should be allocated to mental health and psychological support services to recognize its significance in contributing to treatment success. The advocacy for the utilization of data science and NLP should be encouraged to facilitate informed policy decisions through the insights gained.

Patient advocacy organizations can organize empowerment workshops and support groups that provide patients with the necessary knowledge and skills to actively engage in treatment decisions and self-management. Awareness campaigns can be initiated to educate patients, families and caregivers about the significance of treatment adherence and the range of support options accessible. These organizations also have the potential to collaborate with policymakers in advocating for policies that prioritize patient-centered approaches and utilize their collective efforts to drive broader systemic enhancements.

Implementing these suggestions into the landscape of metastatic breast cancer treatment can foster a patient-centered approach that ultimately elevates patients' adherence to treatment, their outcomes and overall well-being.

Future research

The endeavors of future research could focus on the development of personalized communication strategies that cater to patients' diverse preferences and concerns. Investigating the impact of comprehensive emotional and psychological support on treatment adherence and outcomes would provide valuable insights. Furthermore, exploring innovative technology-driven solutions for continuous real-time adherence monitoring and interventions could significantly enhance patient engagement and persistence. A deeper exploration into the unique challenges faced by patients at different stages of breast cancer, particularly at advanced stages, would yield a nuanced understanding of evolving concerns and needs.

Longitudinal studies tracking patients' adherence behaviors and corresponding clinical outcomes could uncover dynamic patterns and inform targeted interventions. Incorporating patient-generated data, such as online forums and social media discussions, into research methodologies could be a rich source of real-world insights. Analyzing this data using advanced NLP techniques could uncover patients' evolving concerns and thus to shape tailored approaches for effective patient education and support. The investigations into the interplay between patient-centered care, treatment adherence and quality of life would contribute to a holistic understanding of treatment success. Long-term studies assessing the effectiveness of patient-centered interventions on overall treatment outcomes, not just adherence, would provide a more comprehensive evaluation of the impact of these strategies.

Furthermore, as the field of data science and AI continues to evolve, there is a potential to utilize these technologies to develop predictive models that anticipate patients' adherence challenges based on their characteristics and historical data. This proactive approach could enable timely and tailored interventions that ensure sustained treatment engagement.

Additionally, new topic modelling methods have recently evolved which are more focused on the semantic content of the text (BERTopic) [73]. These methods leverage large language models and can also be applied to specific domains using a specific model such as Med-BERT [74].

Finally, collaborations between stakeholders, including healthcare providers, pharmaceutical companies, patient advocacy organizations and policymakers, could yield multi-faceted research initiatives that incorporate clinical,

psychological and societal aspects. Such interdisciplinary research endeavors would contribute to a more holistic understanding of patient-centered care and its profound impact on patient treatment adherence and well-being.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors have no conflicts of interest to declare. All co-authors have read and agreed with the content of the manuscript and there is no financial interest to report. We certify that the submission is an original work and is not under review by another publication.

Data availability

The data are available on request.

Authors' contributions

All authors made substantial contributions to the conception and design of the work; all authors made substantial contributions to the analysis and interpretation of data; drafted the work and revised it critically for important intellectual content; approved the version to be published; agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

[1] Gebbia V, Bellavia G, Ferraù F, Valerio MR. Adherence, compliance and persistence to oral antineoplastic therapy: a review focused on chemotherapeutic and biologic agents. *Expert Opinion on Drug Safety*. 2012;11 Suppl 1:S49-59. doi: 10.1517/14740338.2011.645803.

[2] Partridge AH, Avorn J, Wang PS, Winer EP. Adherence to Therapy With Oral Antineoplastic Agents. *Journal of the National Cancer Institude*. 2002; 94(9):652-61. doi: 10.1093/jnci/94.9.652.

[3] Ruddy K, Mayer E, Partridge A. Patient adherence and persistence with oral anticancer treatment. *CA: A Cancer Journal for Clinicians*. 2009; 59(1): 56-66. doi: 10.3322/caac.20004.

[4] Onwusah DO, Ojewole EB, Chimbari MJ. Adherence to Oral Anticancer Medications Among Women With Breast Cancer in Africa: A Scoping Review. *JCO Global Oncology*. 2023; (9): e2100289. doi: 10.1200/GO.21.00289.

[5] Müller V, Nabieva N, Häberle L, Taran FA, Hartkopf AD, Volz B, et al. Impact of disease progression on health-related quality of life in patients with metastatic breast cancer in the PRAEGNANT breast cancer registry. *Breast*. 2018; 37:154-60. doi: 10.1016/j.breast.2017.08.008.

[6] Chirgwin JH, Giobbie-Hurder A, Coates AS, Price KN, Ejlertsen B, Debled M, et al. Treatment Adherence and Its Impact on Disease-Free Survival in the Breast International Group 1-98 Trial of Tamoxifen and Letrozole, Alone and in Sequence. *Journal of Clinical Oncology*. 2016; 34(21): 2452-9. doi: 10.1200/JCO.2015.63.8619.

[7] World Health Organization. *Breast cancer* [Internet]. 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=In%202020%2C%20there%20were%202.3,the%20world's%20most%20prevalent%20cancer.

[8] Nabieva N, Kellner S, Fehm T, Häberle L, de Waal J, Rezai M, et al. Influence of patient and tumor characteristics on early therapy persistence with letrozole in postmenopausal women with early breast cancer: results of the prospective Evaluate-TM study with 3941 patients. *Annals of Oncology*. 2018; 29(1):186-92. doi: doi.org/10.1093/annonc/mdx630.

[9] Nabieva N, Fehm T, Häberle L, De Waal J, Rezai M, Baier B, et al. Influence of side-effects on early therapy persistence with letrozole in post-menopausal patients with early breast cancer: Results of the prospective EvAluate-TM study. *European Journal of Cancer*. 2018; 96:82-90. doi.org/10.1016/j.ejca.2018.03.020.

[10] Agarwal S, Ying J, Boucher KM, Agarwal JP. The association between socioeconomic factors and breast cancer-specific survival varies by race. *PloS one*. 2017; 12(12): e0187018. doi: 10.1371/journal.pone.0187018.

[11] Ji P, Gong Y, Jiang CC, Hu X, Di GH, Shao ZM. Association between socioeconomic factors at diagnosis and survival in breast cancer: A population-based study. *Cancer Medicine*. 2020; 9(5):1922-36. doi: 10.1002/cam4.2842.

[12] van Maaren MC, Rachet B, Sonke GS, Mauguen A, Rondeau V, Siesling S, et al. Socioeconomic status and its relation with breast cancer recurrence and survival in young women in the Netherlands. *Cancer Epidemiology*. 2022; 77: 102118. doi: 10.1016/j.canep.2022.102118.

[13] Aguado Loi CX, Baldwin JA, McDermott RJ, McMillan S, Martinez Tyson D, Yampolskaya S, et al. Risk factors associated with increased

depressive symptoms among Latinas diagnosed with breast cancer within 5 years of survivorship. *Psychooncology*. 2013; 22(12):2779-88. doi: 10.1002/pon.3357.

[14] daCosta DiBonaventura M, Copher R, Basurto E, Faria C, Lorenzo R. Patient Preferences and Treatment Adherence Among Women Diagnosed with Metastatic Breast Cancer. *American Health & Drug Benefits*. 2014; 7(7): 386-96.

[15] Hurtado-de-Mendoza A, Carrera P, Parrott WG, Gómez-Trillos S, Perera RA, Sheppard VB. Applying the theory of planned behavior to examine adjuvant endocrine therapy adherence intentions. *Psychooncology*. 2019; 28(1):187-94. doi: 10.1002/pon.4931.

[16] Kaptein AA, Schoones JW, van der Meer PB, Matsuda A, Murray M, Heimans L, et al. Psychosocial determinants of adherence with oral anticancer treatment: "we don't need no education." *Acta Oncologica* (Stockholm, Sweden). 2021; 60(1): 87-95. doi: 10.1080/0284186X.2020.1843190.

[17] Lambert LK, Balneaves LG, Howard AF, Gotay CC. Patient-reported factors associated with adherence to adjuvant endocrine therapy after breast cancer: an integrative review. *Breast Cancer Research and Treatment*. 2018;167(3): 615-33. doi: 10.1007/s10549-017-4561-5.

[18] Lambert LK, Balneaves LG, Howard AF, Chia SK, Gotay CC. Understanding adjuvant endocrine therapy persistence in breast Cancer survivors. *BMC Cancer*. 2018;18(1):732. doi: 10.1186/s12885-018-4644-7.

[19] Magai C, Consedine N, Neugut AI, Hershman DL. Common psychosocial factors underlying breast cancer screening and breast cancer treatment adherence: a conceptual review and synthesis. *Journal of Women's Health*. 2007; 16(1): 11-23. doi: 10.1089/jwh.2006.0024.

[20] Chang HA, Barreto N, Davtyan A, Beier E, Cangin MA, Salman J, et al. Depression predicts longitudinal declines in social support among women with newly diagnosed breast cancer. *Psychooncology*. 2019; 28(3): 635-42. doi: 10.1002/pon.5003.

[21] Heisig SR, Shedden-Mora MC, von Blanckenburg P, Rief W, Witzel I, Albert US, et al. What do women with breast cancer expect from their treatment? Correlates of negative treatment expectations about endocrine therapy. *Psychooncology*. 2016; 25(12):1485-92. doi: 10.1002/pon.4089.

[22] Lindberg NM, Wellisch D. Anxiety and compliance among women at high risk for breast cancer. *Annals of Behavioral Medicine: a publication of the Society of Behavioral Medicine*. 2001; 23(4): 298-303. doi: 10.1207/S15324796ABM2304_9.

[23] Pan Y, Heisig SR, von Blanckenburg P, Albert US, Hadji P, Rief W, et al. Facilitating adherence to endocrine therapy in breast cancer: stability and predictive power of treatment expectations in a 2-year prospective study. *Breast Cancer Research and Treatment*. 2018; 168(3): 667-77. doi: 10.1007/s10549-017-4637-2.

[24] Berkowitz MJ, Thompson CK, Zibecchi LT, Lee MK, Streja E, Berkowitz JS, et al. How patients experience endocrine therapy for breast cancer: an online survey of side effects, adherence, and medical team support. *Journal of Cancer Survivorship: research and practice*. 2021; 15(1): 29-39. doi: 10.1007/s11764-020-00908-5.

[25] Conley CC, McIntyre M, Pensak NA, Lynce F, Graham D, Ismail-Khan R, et al. Barriers and facilitators to taking CDK4/6 inhibitors among patients with metastatic breast cancer: a qualitative study. *Breast Cancer Research and Treatment*. 2022; 192(2): 385-99. doi: 10.1007/s10549-022-06518-2.

[26] Engel-Nitz NM, Johnson MG, Johnson MP, Cha-Silva AS, Kurosky SK, Liu X. Palbociclib Adherence and Persistence in Patients with Hormone Receptor Positive/Human Epidermal Growth Factor Receptor 2 Negative (HR+/HER2-) Metastatic Breast Cancer. *Patient Preference and Adherence*. 2023; 17: 1049-62. doi: 10.2147/PPA.S401480.

[27] Fleming L, Agnew S, Peddie N, Crawford M, Dixon D, MacPherson I. The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A quantitative systematic review. *Breast* (Edinburgh, Scotland). 2022; 64: 63-84. doi: 10.1016/j.breast.2022.04.010.

[28] Peddie N, Agnew S, Crawford M, Dixon D, MacPherson I, Fleming L. The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A qualitative systematic review and thematic synthesis. *Breast* (Edinburgh, Scotland). 2021;58:147-59. doi: 10.1016/j.breast.2021.05.005.

[29] Johnson LA. Factors influencing oral adherence: qualitative metasummary and triangulation with quantitative evidence. *Clinical Journal of Oncology Nursing*. 2015; 19(3 Suppl): 6-30. doi: 10.1188/15.S1.CJON.6-30.

[30] Heiney SP, Parker PD, Felder TM, Adams SA, Omofuma OO, Hulett JM. A systematic review of interventions to improve adherence to endocrine therapy. *Breast Cancer Research and Treatment*. 2019; 173(3): 499-510. doi: 10.1007/s10549-018-5012-7.

[31] Mathes T, Antoine SL, Pieper D, Eikermann M. Adherence enhancing interventions for oral anticancer agents: a systematic review. *Cancer Treatment Reviews*. 2014; 40(1): 102-8. doi: 10.1016/j.ctrv.2013.07.004.

[32] Gabitova G, Burke NJ. Improving healthcare empowerment through breast cancer patient navigation: a mixed methods evaluation in a safety-net setting. *BMC Health Services Research*. 2014;

14: 407. doi: 10.1186/1472-6963-14-407.
[33] Heisig SR, Shedden-Mora MC, von Blanckenburg P, Schuricht F, Rief W, Albert US, et al. Informing women with breast cancer about endocrine therapy: effects on knowledge and adherence. *Psychooncology*. 2015; 24(2): 130-7. doi: 10.1002/pon.3611.
[34] Jacobs JM, Pensak NA, Sporn NJ, MacDonald JJ, Lennes IT, Safren SA, et al. Treatment Satisfaction and Adherence to Oral Chemotherapy in Patients With Cancer. *Journal of Oncology Practice*. 2017; 13(5): e474-85. doi: 10.1200/JOP.2016.019729.
[35] Komatsu H, Yagasaki K, Yamaguchi T, Mori A, Kawano H, Minamoto N, et al. Effects of a nurse-led medication self-management programme in women with oral treatments for metastatic breast cancer: A mixed-method randomised controlled trial. *European Journal of Oncology Nursing*. 2020; 47: 101780. doi: 10.1016/j.ejon.2020.101780.
[36] Addeo R, Iodice P, Maiorino L, Febbraro A, Incoronato P, Pisano A, et al. Acceptance and adherence of oral endocrine therapy in women with metastatic breast cancer: exacampania group study. *The Breast Journal*. 2015; 21(3): 326-8. doi: 10.1111/tbj.12409.
[37] Baumann W, Welslau M. Patientenadhärenz in der oralen medizinischen Onkologie. *Der Onkologe*. 2017; 23(8): 645-50. doi: 10.1007/s00761-017-0237-3.
[38] Berry DL, Blonquist TM, Hong F, Halpenny B, Partridge AH. Self-reported adherence to oral cancer therapy: relationships with symptom distress, depression, and personal characteristics. *Patient Preference and Adherence*. 2015; 9: 1587-92. doi: 10.2147/PPA.S91534.
[39] Chlebowski RT, Geller ML. Adherence to endocrine therapy for breast cancer. *Oncology*. 2006; 71(1-2):1-9. doi: 10.1159/000100444.
[40] Doggrell SA. Adherence to oral endocrine treatments in women with breast cancer: can it be improved? *Breast Cancer Research and Treatment*. 2011; 129(2): 299-308. doi: 10.1007/s10549-011-1578-z.
[41] Geynisman DM, Wickersham KE. Adherence to targeted oral anticancer medications. *Discovery Medicine*. 2013; 15(83): 231-41.
[42] Gillespie TW. Advances in Oral Oncolytic Agents for Breast Cancer and Recommendations for Promoting Adherence. *Journal of the Advanced Practitioner in Oncology*. 2020; 11(1): 83-96. doi: 10.6004/jadpro.2020.11.1.5.
[43] Gotay C, Dunn J. Adherence to long-term adjuvant hormonal therapy for breast cancer. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2011; 11(6): 709-15. doi: 10.1586/erp.11.80.
[44] Huang WC, Chen CY, Lin SJ, Chang CS. Medication adherence to oral anticancer drugs: systematic review. *Expert Review of Anticancer Therapy*. 2016; 16(4): 423-32. doi: 10.1586/14737140.2016.1159515.
[45] Jacobs JM, Ream ME, Pensak N, Nisotel LE, Fishbein JN, MacDonald JJ, et al. Patient Experiences With Oral Chemotherapy: Adherence, Symptoms, and Quality of Life. *Journal of the National Comprehensive Cancer Network: JNCCN*. 2019; 17(3): 221-8. doi: 10.6004/jnccn.2018.7098.
[46] Kimmick G, Edmond SN, Bosworth HB, Peppercorn J, Marcom PK, Blackwell K, et al. Medication taking behaviors among breast cancer patients on adjuvant endocrine therapy. *Breast* (Edinburgh, Scotland). 2015; 24(5): 630-6. doi: 10.1016/j.breast.2015.06.010.
[47] Lin C, Clark R, Tu P, Bosworth HB, Zullig LL. Breast cancer oral anti-cancer medication adherence: a systematic review of psychosocial motivators and barriers. *Breast Cancer Research and Treatment*. 2017; 165(2): 247-60. doi: 10.1007/s10549-017-4317-2.
[48] Lin ID, Shotts MB, Al-Hader A, Weddle KJ, Holden RJ, Mueller EL, et al. Examining adherence to oral anticancer medications through a human factors engineering framework: Protocol for a scoping review. *PloS One*. 2022;17(9): e0274963. doi: 10.1371/journal.pone.0274963.
[49] Milata JL, Otte JL, Carpenter JS. Oral Endocrine Therapy Nonadherence, Adverse Effects, Decisional Support, and Decisional Needs in Women With Breast Cancer. *Cancer Nursing*. 2018; 41(1): E9-18. doi: 10.1097/NCC.0000000000000430.
[50] Mislang AR, Wildes TM, Kanesvaran R, Baldini C, Holmes HM, Nightingale G, et al. Adherence to oral cancer therapy in older adults: The International Society of Geriatric Oncology (SIOG) taskforce recommendations. *Cancer Treatment Reviews*. 2017; 57: 58-66. doi: 10.1016/j.ctrv.2017.05.002.
[51] Verbrugghe M, Timmers L, Boons CCLM, van den Bemt BJF, Hugtenburg JG, van Hecke A. Adherence to oral anticancer agents: Healthcare providers' perceptions, beliefs and shared decision making in Belgium and the Netherlands. *Acta Oncologica* (Stockholm, Sweden). 2016; 55(4): 437-43. doi: 10.3109/0284186X.2015.1119307.
[52] Milic M, Foster A, Rihawi K, Anthoney A, Twelves C. "Tablet burden" in patients with metastatic breast cancer. *European Journal of Cancer* (Oxford, England: 1990). 2016; 55: 1-6. doi: 10.1016/j.ejca.2015.11.015.
[53] Neugut AI, Zhong X, Wright JD, Accordino M, Yang J, Hershman DL. Nonadherence to Medications for Chronic Conditions and Nonadherence to Adjuvant Hormonal Therapy in Women With Breast Cancer. *JAMA Oncology*. 2016; 2(10): 1326-32. doi:10.1001/jamaoncol.2016.1291.
[54] Paranjpe R, John G, Trivedi M, Abughosh S. Identifying adherence barriers to oral endocrine therapy among breast cancer survivors. *Breast*

Cancer Research and Treatment. 2019; 174(2): 297-305. doi: 10.1007/s10549-018-05073-z.

[55] Yang J, Neugut AI, Wright JD, Accordino M, Hershman DL. Nonadherence to Oral Medications for Chronic Conditions in Breast Cancer Survivors. *Journal of Oncology Practice*. 2016; 12(8): e800-9. doi: 10.1200/JOP.2016.011742.

[56] Muluneh B, Deal A, Alexander MD, Keisler MD, Markey JM, Neal JM, et al. Patient perspectives on the barriers associated with medication adherence to oral chemotherapy. *Journal of Oncology Pharmacy Practice*. 2018; 24(2): 98-109. doi: 10.1177/1078155216679026.

[57] Salgado TM, Mackler E, Severson JA, Lindsay J, Batra P, Petersen L, et al. The relationship between patient activation, confidence to self-manage side effects, and adherence to oral oncolytics: a pilot study with Michigan oncology practices. *Supportive Care in Cancer*. 2017; 25(6): 1797-807. doi: 10.1007/s00520-017-3584-0.

[58] Stanton AL, Petrie KJ, Partridge AH. Contributors to nonadherence and nonpersistence with endocrine therapy in breast cancer survivors recruited from an online research registry. *Breast Cancer Research and Treatment*. 2014; 145(2): 525-34. doi: 10.1007/s10549-014-2961-3.

[59] Stilley CS, Bender CM, Dunbar-Jacob J, Sereika S, Ryan CM. The impact of cognitive function on medication management: three studies. *Health Psychology*. 2010; 29(1): 50-5. doi: 10.1037/a0016940.

[60] Timmers L, Boons CCLM, Verbrugghe M, van den Bemt BJF, van Hecke A, Hugtenburg JG. Supporting adherence to oral anticancer agents: clinical practice and clues to improve care provided by physicians, nurse practitioners, nurses and pharmacists. *BMC Cancer*. 2017; 17(1): 122. doi: 10.1186/s12885-017-3110-2.

[61] Wood L. A review on adherence management in patients on oral cancer therapies. *European Journal of Oncology Nursing*. 2012; 16(4): 432-8. doi: 10.1016/j.ejon.2011.10.002.

[62] World Health Organization. *Adherence to long-term therapies: Evidence for action*. Geneva: World Health Organization; 2003.

[63] Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. *Medicine and Pharmacy Reports*. 2019; 92(2):117-22. doi: 10.15386/mpr-1201.

[64] Levit LA, Arora S, Kluetz PG, Magnuson A, Rahman A, Harvey RD. Call to Action for Improving Oral Anticancer Agent Adherence. *Journal of Clinical Oncology*. 2022; 40(10): 1036-40. doi: 10.1200/JCO.21.02529.

[65] DiMatteo MR. Variations in Patients' Adherence to Medical Recommendations: A Quantitative Review of 50 Years of Research. *Medical Care*. 2004; 42(3): 200-9. doi: 10.1097/01.mlr.0000114908.90348.f9.

[66] Bluethmann S, Murphy C, Tiro J, Mollica M, Vernon S, Bartholomew LK. Deconstructing Decisions to Initiate, Maintain, or Discontinue Adjuvant Endocrine Therapy in Breast Cancer Survivors: A Mixed-Methods Study. *Ocology Nursing Forum*. 2017; 44(3): E101-10. doi: 10.1188/17.ONF.E101-E110.

[67] Albrecht J, Ramachandran S, Winkler C. *Blueprints for text analysis using Python: Machine learning-based solutions for common real world (NLP) applications* [Internet]. Sebastopol, CA: O'Reilly; 2021. Available from: https://www.oreilly.com/library/view/blueprints-for-text/9781492074076/.

[68] Blei DM. Probabilistic topic models. *Communications of the ACM*. 2012; 55(4): 77-84. doi/10.1145/2133806.2133826.

[69] Foong NW. *Introduction to Topic Modeling using Scikit-Learn: Explore 3 unsupervised techniques to extract important topics from documents* [Internet]. Towards Data Science, editor. 2021. Available from: https://towardsdatascience.com/introduction-to-topic-modeling-using-scikit-learn-4c3f3290f5b9.

[70] Arroyo-Fernández I, Méndez-Cruz CF, Sierra G, Torres-Moreno JM, Sidorov G. Unsupervised sentence representations as word information series: Revisiting TF-IDF. *Computer Speech & Language*. 2019; 56: 107-29. doi:10.1016/j.csl.2019.01.005.

[71] Aladakatti SS, Senthil Kumar S. Exploring natural language processing techniques to extract semantics from unstructured dataset which will aid in effective semantic interlinking. *International Journal of Modeling, Simulation, and Scientific Computing*. 2023; 14(01). doi: 10.1142/S1793962322430048.

[72] Anantrasirichai N, Bull D. Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*. 2022; 55(1): 589-656. doi: 10.1007/s10462-021-10039-7.

[73] Grootendorst M. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* [Internet]. arXiv; 2022 [cited 2024 Jul 31]. Available from: http://arxiv.org/abs/2203.05794.

[74] Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*. 2021; 4(1): 86. doi: 10.1038/s41746-021-00455-y.